DOCUMENT RESUME

ED 132 180 95 TM 005 895

AUTHOR Everett, Bruce E.
TITLE A Preliminary Study of the Relevance of a Standardized Test for Measuring Achievement Gains in Innovative Arithmetic Programs. Project LONGSTEP Final Report. Volume II. Appendix Report.
INSTITUTION American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO AIR-21400-9/76-FR-II (App)
PUB DATE Sep 76
CONTRACT OEC-0-70-4789
NOTE 20p.; For related documents, see TM 005 891-896 and TM 005 920-922

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Achievement Tests; *Arithmetic; *Educational Innovation; Educational Objectives; Elementary School Students; Elementary Secondary Education; *Longitudinal Studies; Relevance (Education); *Standardized Tests
IDENTIFIERS *Comprehensive Tests of Basic Skills; Longitudinal Study of Educational Practices; *Project LONGSTEP

ABSTRACT

The general emphasis of Project LONGSTEP was on the identification of changes in student achievement that occur as a result of exposure to intensive educational innovation. During the course of the project, questions were raised concerning (1) the relevance of the standardized achievement test utilized in the project to the stated objectives of the educational treatments included in the participating educational programs and (2) the extent to which the findings may have been influenced by the particular instrument chosen to measure cognitive achievement. To provide at least partial answers to these questions, a study of the arithmetic items contained in the test and their relevance to the educational objectives of the arithmetic treatments encompassed by the project was conducted. Specifically, two research questions were addressed: (1) Were the particular skills necessary to answer correctly the arithmetic items on the Comprehensive Tests of Basic Skills (CTBS) actually incorporated into the curriculum objectives of arithmetic treatments? and (2) To the extent that arithmetic treatments differed from one another in the relevance of the CTBS test items to their curriculum objectives, what effect did this difference have on test performance? "Data Collection Instruments and Guidelines" developed for Project LONGSTEP referenced in Vol. I, Chapter II, Section C, will be accessioned TM 005 987 in RIEMAY77. (RC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED132180

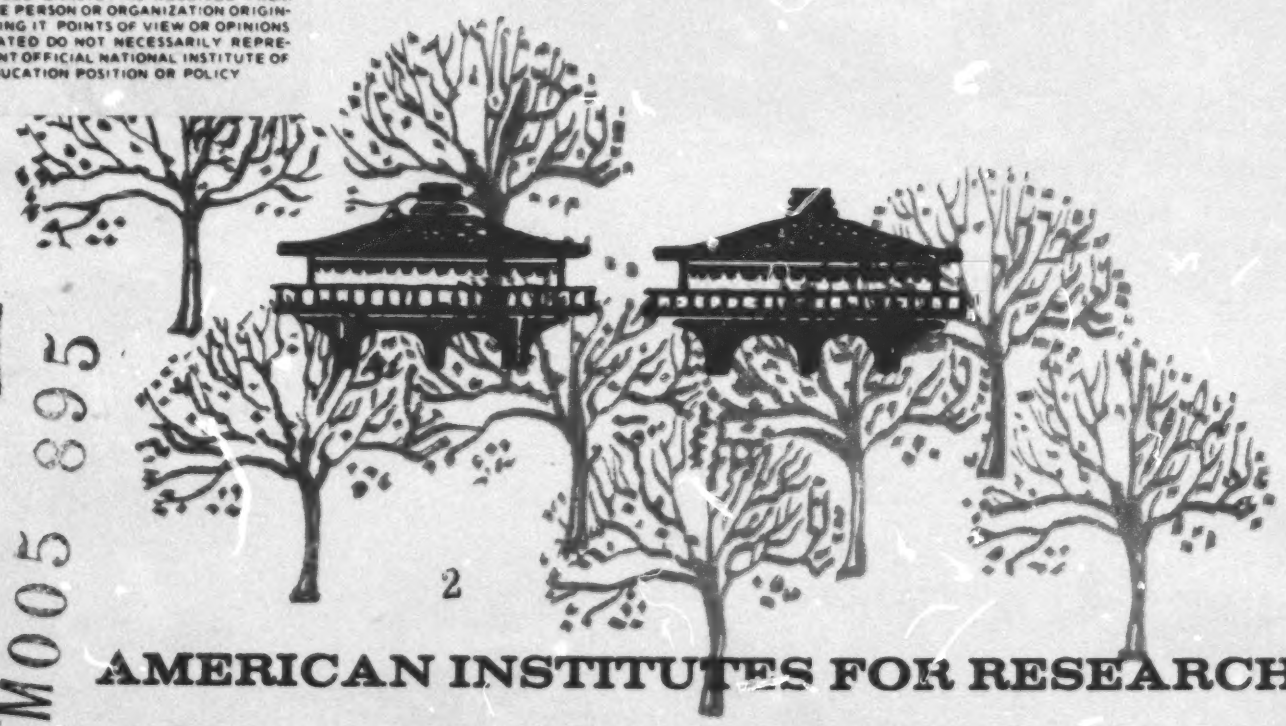
A PRELIMINARY STUDY OF THE RELEVANCE OF A STANDARDIZED TEST FOR MEASURING ACHIEVEMENT GAINS IN INNOVATIVE ARITHMETIC PROGRAMS

Project LONGSTEP Final Report: Volume II

APPENDIX REPORT

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY



2

AMERICAN INSTITUTES FOR RESEARCH

Palo Alto, California

TM005 895

AIR-21400-9/76-FR-II (App)

A PRELIMINARY STUDY OF THE RELEVANCE OF A STANDARDIZED TEST FOR MEASURING ACHIEVEMENT GAINS IN INNOVATIVE ARITHMETIC PROGRAMS

Bruce S. Everett

PROJECT LONGSTEP FINAL REPORT: VOLUME II

APPENDIX REPORT

Prepared for

Office of Planning, Budgeting and Evaluation
Office of Education
Washington, D. C. 20202

American Institutes for Research
Palo Alto, California 94302

September 1976

The research reported herein was performed pursuant to Contract No. OEC-0-70-4789 with the U. S. Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

PREFACE

This document is an Appendix Report to Volume II of the final report of Project LONGSTEP, the Longitudinal Study of Educational Practices. It addresses the topic of the relevance of a standardized mathematics test for measuring achievement gains for third- and sixth-graders participating in the last year of the study.

We would like to acknowledge the help of Doris A. Alfred, Standish W. Sibley, and David R. Gross in preparing the data used in this report.

Introduction

Specific objectives of Project LONGSTEP were to design a system to study the characteristics underlying innovative educational approaches; to establish a large-scale data base of program characteristics and student outcomes for a select sample of educational programs involving intensive and highly innovative education practices; to determine longitudinally the impact of such innovation upon student performance and attitudes; and to attempt to identify the dimensions of the components that exhibited the greatest impact on student outcomes. A complete description of Project LONGSTEP design and data collection methods can be found in Volume I of the final report (Coles, Chalupsky, Everett, Shaycoft, Rodabaugh and Danoff, 1976).

During the course of Project LONGSTEP questions were raised concerning (a) the relevance of the standardized achievement test utilized in the project to the stated objectives of the educational treatments included in the participating educational programs, and (b) the extent to which the findings discussed in Volume I may have been influenced by the particular instrument chosen to measure cognitive achievement. To provide at least partial answers to these questions, a study was conducted of the arithmetic items contained in the test and their relevance to the educational objectives of the arithmetic treatments encompassed by Project LONGSTEP. The findings of this study are presented in this appendix.

Objective of This Study

The objective of this study was to investigate the extent to which the stated curriculum objectives of the arithmetic treatments included in Project LONGSTEP were associated with the performance of students on the arithmetic section of the CTBS.[1] To this end the following two research

[1] Comprehensive Tests of Basic Skills (CTBS), 1968 edition, Monterey, California, CTB/McGraw-Hill.

questions were addressed:

- Were the particular skills necessary to answer arithmetic items on the CTBS correctly actually incorporated into the the curriculum objectives of arithmetic treatments?
- To the extent that arithmetic treatments differed from one another in the relevance of the CTBS test items to their curriculum objectives, what effect did this difference have on test performance?

As part of the data collection for the 1973-74 school year, site visitors gathered information on the arithmetic objectives of selected treatments documented during the 1972-73 school year. The original purpose of this effort was to explore the question of whether the adjustment of a standardized achievement test to better match individual educational treatment objectives would result in substantially different findings from those obtained by the use of total test scores.

Study Procedure

It was initially thought that there would be some merit in "adjusting" student test scores in such a way that they would be consistent with the objectives of their own educational treatments. Unfortunately, direct adjustment of the CTBS scores would have the effect of turning that test into a criterion-referenced instrument; a given score would only have meaning within the context of a particular treatment. Without a consistent criterion measure, any assessment of differences among treatments would be virtually impossible. In order to develop consistent criterion measures, it was decided to separate the arithmetic items on the CTBS into item domains, that is, sets of items which were highly interrelated with one another in terms of required skills and less related to items requiring different skills.

Similarly, the curriculum objectives of each arithmetic treatment were separated into sets of interrelated skills which paralleled the aforementioned item domains. Although the curriculum objectives of these treatments

might have covered content areas other than those tapped by the CTBS, the absence of a criterion measure which addressed such topics precluded any assessment of student performance on them. Of the curriculum objectives that were tapped by the CTBS, however, it was possible to match them to actual student performance, that is, scores within specified item domains. A further restriction was that only item domains which had at least five items would be considered for analysis. The indicator actually used to assess each student's achievement within a domain was the number of items within the domain answered correctly.

The two target populations of students chosen to test out these ideas were the third-graders and the sixth-graders who participated in the final year of Project LONGSTEP's data collection (1972-73). Both grades had received the same level of the CTBS in the previous, or pretest, year as they had in the target, or posttest, year.

In order to obtain a precise assessment of the curriculum objectives of the educational treatments to which these target students belonged, a list of the specific skills associated with each item on the CTBS Arithmetic Test, Levels 1 and 2, was prepared.[2] These lists were given to the third and sixth grade teachers of the target students with the instructions to check each of the listed skills which had been taught to most of their students during the target year. A particular skill was considered relevant to a treatment's curriculum objectives if a majority of the teachers in a treatment said that they had taught that skill to their students during the target year. In effect, curriculum objectives were empirically determined for each treatment through the use of these lists; the total number of appropriate objectives within a given content area constituted a measure of the relevance of the content area to the domain for each treatment.

For each educational treatment, then, there were to be two types of measures of specific arithmetic skills: (1) the pretest and posttest scores of the students in the treatment on those items on the CTBS which related to a particular content area, and (2) the assessment of the teacher

[2]Appreciation is expressed to Dr. William Shanner, who provided project staff with an item-by-item list of the objectives contained in the CTBS.

or teachers in that treatment as to whether those students could be expected to have encountered that content area during the past school year. For each content area, there would be an item domain and a parallel measure of domain relevance. If an item domain were relevant, one would expect higher posttest scores than if the domain were not considered relevant.

The four item domains thus chosen for the third grade sample were: Simple Arithmetic Computation (add two 2-digit numbers without regrouping), Simple Arithmetic Word Problems (solve a story problem that requires multiplication by a 1-digit multiplier), Units of Measurement (given linear measures of feet or inches, express one measure in terms of the other), and Numerical Relationships (recognize one number as being greater than, equal to, or less than a second number). The eight item domains chosen for the sixth grade sample were: Simple Arithmetic Computation (multiply any 3-digit number by a 2-digit number), Use of Decimals (subtract decimals through hundred-thousandths), Simple Fractions (add unlike fractions and express the number in lowest terms), Compound Fractions (divide a whole number by a mixed number or vice-versa), Units of Measurement (add or subtract measures of time, giving the answer in simplest form), Simple Arithmetic Word Problems (solve a two-step word problem involving multiplication and addition), Fractions in Word Problems (solve a word problem involving addition or subtraction of fractions and express the answer in lowest terms), and Numerical Relationships (given four integers in a series, determine the next in the series). A composite total of the selected domains was also created.

Three additional types of data on each student were also included in the analysis: a scale of socioeconomic status for each student, the Number of Minutes per Day spent in arithmetic instruction, and the overall Level of Innovation scale for each student's arithmetic treatment group. The final analytic model consisted of indices of the following: (1) student SES, (2) student pretest scores for each domain, (3) the Number of Minutes per Day spent in arithmetic instruction, (4) Level of Innovation of each student's treatment, (5) the degree of relevance of each domain to the objectives of the student's arithmetic curriculum, and (6) student posttest scores for each domain.

Results

The findings of our survey of arithmetic treatment objectives are summarized in Table 1. Of the four arithmetic-item domains selected for analysis at the third grade level, all of them were considered to be relevant to the curriculum objectives of at least 76% of the arithmetic treatments that were sampled. By implication, 97 of the 110 items of the CTBS Level 1, Form Q Arithmetic Section were considered to be within the grasp of the students included in three-fourths of the arithmetic treatments studied. Of the eight arithmetic-item domains selected for analysis at the sixth grade level, five were reported as relevant to the curriculum objectives of at least 95% of the arithmetic treatments that were sampled. This would imply that 67 of the 98 items of the CTBS Level 2, Form Q Arithmetic Section were considered to be relevant for nearly all of these students.

A comparison of the number of objectives and items associated with each domain shows a strong emphasis on simple arithmetic concepts at both levels of the CTBS. Not only do simple arithmetic computation problems and simple arithmetic word problems make up a large proportion of the items, but they also tend to involve a limited number of concepts. Objectives required for other domains may be more numerous, but they count less toward the overall arithmetic total. Moreover, items not included in this analysis because of their limited contribution to the CTBS tended to require skills that were well above grade level (geometry) or were not strictly arithmetic in content (use of metric terms).

The stepwise multiple regressions of item domain posttest scores on Number of Minutes per Day in arithmetic, Level of Innovation in arithmetic, Student SES, item domain pretest, and item domain relevance for the third grade sample are presented in Table 2. Domain Relevance uniquely accounts for 1.3% of the variance in Simple Arithmetic Computation and accounts for even less of the variance in posttest performance regarding the other item domains. Pretest consistently shows the largest beta weight for all four domains, and is especially important in its contribution to posttest performance on the composite total of selected items. Level of Innovation shows a consistently significant negative association with posttest, whereas

TABLE 1

Relevance of Treatment Objectives to Selected CTBS Item Domain

Domain	No. of Treatments for Whom the Domain is Relevant	% of Treatments For Whom the Domain is Relevant	No. of Objectives	No. of Items on CTBS
Third Grade - 17 Different Treatments				
Simple Arithmetic Computation	14	82%	19	59
Simple Arithmetic Word Problems	16	94%	7	24
Units of Measurement	13	76%	4	8
Numerical Relationships	15	88%	6	6
Composite Total	13	76%	36	97
Sixth Grade - 37 Different Treatments				
Simple Arithmetic Computations	37	100%	9	26
Use of Decimals	35	95%	10	10
Simple Fractions	22	59%	2	5
Complex Fractions	35	95%	10	7
Simple Arithmetic Word Problems	37	100%	6	14
Fraction Word Problems	19	51%	3	6
Units of Measurement	36	97%	6	10
Numerical Relationships	7	19%	1	7
Composite Total	36	97%	47	85

TABLE 2

Stepwise Multiple Correlation of Selected Predictors with Student Item Domain Scores

Grade 3 (df = 5,499)

Variable	R^2	B	Beta	F
Domain 1 - Simple Arithmetic Computation:				
Number of Minutes per Day in arithmetic	0.009	0.06	0.05	1.50
Level of Innovation in arithmetic	0.049	-0.68	-0.23	32.57**
Student SES	0.084	0.14	0.14	11.94**
Pretest	0.168	0.37	0.29	46.92**
Domain Relevance	0.181	0.38	0.12	7.86**
(Constant)		29.15		
Domain 2 - Simple Arithmetic Word Problems:				
Number of Minutes per Day in arithmetic	0.021	0.04	0.10	5.97*
Level of Innovation in arithmetic	0.045	-0.17	-0.19	17.92**
Student SES	0.102	0.06	0.19	18.27**
Pretest	0.157	0.21	0.24	32.23**
Domain Relevance	0.157	0.00	0.00	0.00
(Constant)		0.99		

*p = ≤ .05
**p = ≤ .01

(continued)

TABLE 2 (continued)

Variable	R^2	B	Beta	F
Domain 3 - Units of Measurement:				
Number of Minutes per Day in arithmetic	0.002	0.00	0.02	0.12
Level of Innovation in arithmetic	0.008	-0.03	-0.09	4.59*
Student SES	0.040	0.02	0.13	8.62**
Pretest	0.068	0.18	0.18	15.99**
Domain Relevance	0.069	-0.05	-0.04	0.65
(Constant)		1.63		
Domain 4 - Numerical Relationships:				
Number of Minutes per Day in arithmetic	0.019	0.02	0.10	4.60*
Level of Innovation in arithmetic	0.042	-0.09	-0.18	18.30**
Student SES	0.083	0.03	0.17	14.40**
Pretest	0.124	0.30	0.21	23.39**
Domain Relevance	0.124	-0.06	-0.02	0.20
(Constant)		1.69		

*p = $\leq$.05
**p = $\leq$.01

(continued)

TABLE 2 (continued)

Variable	R^2	B	Beta	F
Composite Total of Selected Item Domains:				
Number of Minutes per Day in arithmetic	0.010	-0.05	-0.01	0.12
Level of Innovation in arithmetic	0.048	-2.82	-0.23	60.08**
Student SES	0.100	0.02	0.01	0.03
Pretest	0.561	0.78	0.72	519.41**
Domain Relevance	0.563	0.36	0.04	1.78
(Constant)		174.66		

*p = ≤.05
**p = ≤.01

Number of Minutes per Day spent in arithmetic tends to have a slight positive effect. SES shows a small positive effect on each domain posttest, but this effect does not appear with respect to the composite total score on selected item domains for the third grade sample.

The stepwise multiple regressions of item domain posttest scores on the selected predictors for the sixth grade are presented in Table 3. Domain relevance has a significant association with posttest performance on Compound Fractions, Simple Arithmetic Word Problems and the Composite Domain Total. However, because of the number of students involved, these findings are misleading; the percent of posttest variance actually explained by domain relevance is less than 1% in all cases.

Pretest is far and away the most powerful predictor of sixth grade posttest performance in this model, and it tends to overshadow the small but consistently positive effect on item domain scores due to SES. Level of Innovation in arithmetic has a consistent negative association with posttest performance on all eight domains, but is significantly associated only with performance regarding Use of Decimals and Compound Fractions. The Number of Minutes per Day spent in arithmetic is significant in its impact on posttest performance on Simple Fractions, Compound Fractions, and the Composite Domain Total but does not show much contribution to scores in the other item domains.

Discussion

With respect to our first research question, whether or not the arithmetic skills tapped by the CTBS were actually incorporated into curriculum objectives of arithmetic treatments, the answer is clearly "Yes" for the third and sixth grade samples selected from Project LONGSTEP. The CTBS does appear to be a valid measure of the degree to which student performance in arithmetic matches the arithmetic objectives of these particular treatments. Although we do not know how widely these treatments differ among themselves in terms of objectives that do not appear on the CTBS, there are very few grade level arithmetic skills which are not present in some way on the CTBS. What disagreement there is about the relevance of

TABLE 3

Stepwise Multiple Correlation of Selected Predictors with Student Item Domain Scores

Grade 6 (df = 5,1162)

Variable	R^2	B	Beta	F
Domain 1 - Simple Arithmetic Computation:				
Number of Minutes per Day in arithmetic	0.001	0.01	0.02	0.50
Level of Innovation in arithmetic	0.016	-0.03	-0.04	2.22
Student SES	0.048	0.02	0.06	5.31*
Pretest	0.347	0.43	0.56	534.80**
Domain Relevance	0.349	0.12	0.04	2.48
(Constant)		11.35		
Domain 2 - Use of Decimals:				
Number of Minutes per Day in arithmetic	0.001	0.01	0.03	1.38
Level of Innovation in arithmetic	0.024	-0.04	-0.07	6.87**
Student SES	0.067	0.02	0.13	21.87**
Pretest	0.283	0.42	0.48	345.64**
Domain Relevance	0.283	-0.01	-0.01	0.10
(Constant)		2.98		

*p = $\leq$ .05
**p = $\leq$ .01

(continued)

TABLE 3 (continued)

Variable	R^2	B	Beta	F
Domain 3 - Simple Fractions:				
Number of Minutes per Day in arithmetic	0.002	0.01	0.06	4.45*
Level of Innovation in arithmetic	0.023	-0.02	-0.05	2.93
Student SES	0.062	0.02	0.13	19.90**
Pretest	0.189	0.33	0.36	180.09**
Domain Relevance	0.191	0.12	0.05	2.73
(Constant)		0.38		
Domain 4 - Compound Fractions:				
Number of Minutes per Day in arithmetic	0.004	0.02	0.10	14.40**
Level of Innovation in arithmetic	0.058	-0.08	-0.15	31.38**
Student SES	0.109	0.03	0.14	25.97**
Pretest	0.235	0.37	0.36	186.54**
Domain Relevance	0.242	0.13	0.08	9.99**
(Constant)		-0.31		

*p = ≤ .05
**p = ≤ .01

(continued)

TABLE 3 (continued)

Variable	R^2	B	Beta	F
Domain 5 - Units of Measurement:				
Number of Minutes per Day in arithmetic	0.000	0.01	0.03	2.18
Level of Innovation in arithmetic	0.022	-0.02	-0.03	1.37
Student SES	0.036	0.03	0.09	17.38**
Pretest	0.473	0.65	0.65	841.44**
Domain Relevance	0.474	-0.34	-0.03	2.30
(Constant)		3.88		
Domain 6 - Simple Arithmetic Word Problems:				
Number of Minutes per Day in arithmetic	0.001	0.00	0.00	0.00
Level of Innovation in arithmetic	0.032	-0.01	-0.01	0.30
Student SES	0.113	0.02	0.14	30.91**
Pretest	0.367	0.55	0.52	434.34**
Domain Relevance	0.376	0.19	0.11	17.55**
(Constant)		-1.07		

*p = ≤ .05
**p = ≤ .01

(continued)

TABLE 3 (continued)

Variable	R^2	B	Beta	F
Domain 7 - Fractions in Word Problems:				
Number of Minutes per Day in arithmetic	0.008	0.00	0.00	0.01
Level of Innovation in arithmetic	0.028	-0.02	-0.04	2.93
Student SES	0.116	0.03	0.18	45.57**
Pretest	0.347	0.48	0.50	405.33**
Domain Relevance	0.347	0.03	0.01	0.10
(Constant)		0.37		
Domain 8 - Numerical Relationships:				
Number of Minutes per Day in arithmetic	0.002	0.00	-0.01	0.08
Level of Innovation in arithmetic	0.016	-0.01	-0.04	1.79
Student SES	0.054	0.01	0.08	8.78**
Pretest	0.297	0.44	0.51	399.97**
Domain Relevance	0.297	0.04	0.01	0.22
(Constant)		2.69		
Composite Total of Selected Item Domains:				
Number of Minutes per Day in arithmetic	0.001	0.26	0.03	3.95*
Level of Innovation in arithmetic	0.039	-0.35	-0.02	1.17
Student SES	0.133	0.37	0.06	8.63**
Pretest	0.661	0.87	0.78	1774.78**
Domain Relevance	0.662	0.56	0.04	4.06*
(Constant)		30.04		

*p = ≤.05
**p = ≤.01

CTBS items among the teachers polled is largely confined to the more complex, difficult objectives, and these objectives are tapped by a much smaller number of items than are those relating to basic arithmetic skills.

In response to the second research question, to what extent variations in the relevance of the CTBS to curriculum objectives affect test performance, the fact that there is so little variance among treatments in the relevance of the CTBS probably accounts for the finding that the impact of domain relevance is so slight. Also, the pretest scores for the two samples show themselves to be much more important predictors of posttest performance than does domain relevance. The importance of current exposure to these arithmetic skills thus appears to be largely overshadowed by prior experience with those skills.

In general, the results in this study closely parallel Project LONGSTEP's overall findings. Pretest is the predominant predictor of posttest, especially with regard to the more elementary item domains, which are the very ones that students would be expected to have encountered in the previous school year. The more complex domains are less affected by pretest, but they also make up a smaller proportion of the CTBS Arithmetic Total.

Another finding which is consistent with Project LONGSTEP's overall findings is the apparent negative effect of innovative practices on student performance in the last year of the study, especially in grade 3. This may be related to a general "tightening up" of educational practices observed over the course of the study, but it may also be related to the ineffectiveness of innovative techniques in transmitting basic arithmetic concepts to students.

REFERENCES

Coles, G. J., Chalupsky, A. B., Everett, B. E., Shaycoft, M. F., Rodabaugh, B. J., & Danoff, M. N. Impact of educational innovation on student performance: Project methods and findings for three cohorts. Project LONGSTEP Final Report: Volume I. Palo Alto, California: American Institutes for Research, 1976.